DOCUMENT RESUME

ED 126 143 TM 005 399

AUTHOR Martin, Charles G.; Games, Paul A.
TITLE ANOVA Tests of Homogeneity of Variance When n's Are Unequal.
PUB DATE [Apr 76]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Comparative Analysis; Error Patterns; Hypothesis Testing; Measurement Techniques; *Sampling; Simulation; *Tests of Significance
IDENTIFIERS *Box Test; Homogeneity of Variance; *Jackknife; Power (Statistics); Type I Error

ABSTRACT

Stability of Type I error rates and power are investigated for three forms of the Box test and two forms of the jackknife test with equal and unequal sample sizes under conditions of normality and nonnormality. The Box test is shown to be robust to violations of the assumption of normality when sampling is from leptokurtic populations. The jackknife test is not robust. When n's are unequal, previously reported suggestions for selecting subsample sizes for the Box test are shown to be inappropriate, producing a permissive test. The problem of heterogeneous within cell variance and unequal n's is discussed. Two procedures which alleviate this problem are presented for the Box test. Use of the jackknife test with a reduced alpha is shown to provide power and control of Type I error at approximately the same level as the Box test. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

23.22

ED126143

ANOVA Tests of Homogeneity of Variance When n's Are Unequal

Charles G. Martin and Paul A. Games*

The Pennsylvania State University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 399

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, Calif., April 1976

ABSTRACT

Stability of Type I error rates and power are investigated for three forms of the Box test and two forms of the jackknife test with equal and unequal sample sizes under conditions of normality and nonnormality. The Box test is shown to be robust to violations of the assumption of normality when sampling is from leptokurtic populations. The jackknife test is not robust. When n's are unequal, previously reported suggestions for selecting subsample sizes for the Box test are shown to be inappropriate, producing a permissive test. The problem of heterogeneous within cell variances and unequal n's is discussed. Two procedures which alleviate this problem are presented for the Box test. Use of the jackknife test with a reduced alpha is shown to provide power and control of Type I error at approximately the same level as the Box test.

1. INTRODUCTION

In research, it frequently happens that populations are to be tested for homogeneity of variance in complex, multi-factor experimental designs. Methods of statistical inference concerning variances generally have been restricted to single factor designs. Typically, independent random samples are compared via some function of the sample variances with a known sampling distribution when assumptions are met. Many of these tests on variances are notoriously sensitive to distribution form, particularly to kurtosis. Martin and Games (1975) provided a review of many of these tests.

Several researchers have applied the analysis of variance (AOV) to testing variances in order to take advantage of its robustness to non-normality. These techniques have the added benefit of being generalizable to complex, multi-factor experimental designs. A multiplicative model for testing variances, shown to be analogous to the additive AOV model for means through the logarithmic transformation, was presented by Bechhofer (1960). A variety of multi-factor AOV designs for testing equality of variances which fit the model were also presented.

Bartlett and Kendall (1946) suggested performing an AOV on logarithmic transformations of variance estimates. Box (1953) suggested dividing samples into subsamples and performing an AOV on logarithmic transformations of variance estimates computed on the subsamples. The technique was generalized by Scheffé (1959, p. 83) to accommodate unequal subsample sizes and sampling from nonnormal populations. This statistic is

described in Winer (1971, p. 219). A modification of this test to accommodate unequal sample sizes due to Bargmann was presented by Gartside (1972). Levene (1960) proposed performing an AOV on $Z_i = |X_i - \overline{X}|$ or $S_i = (X_i - \overline{X})^2$. The Levene Z statistic is recommended as a test of homogeneity of variance by Glass and Stanley (1970, p. 506). An alternative Levene statistic in which absolute deviations from the mean are replaced by absolute deviations from the sample median, $Z_i' = |X_i - \text{median}|$, was suggested by Miller (1968). Brown and Forsythe (1974) suggested a Levene statistic in which absolute deviations are taken from the ten percent trimmed mean, the mean after deleting the ten percent largest and ten percent smallest values in that group. Miller (1968) also suggested jackknifing logarithms of variance estimates as a test of homogeneity of variance. The Z-variance statistic, an AOV test utilizing the z-score transformation for chi-square statistics based on large degrees of freedom, was developed by Overall and Woodward (1974).

The Levene Z statistic was shown to be sensitive to distribution form with inflated P(EI)'s by Brown and Forsythe (1974), Fellers (1972), and Games, Winkler and Probert (1972). Miller (1968) proved that the Z test is not asymptotically distribution free. Neel and Stallings (1974) showed that even with a normal population, when n's are small the Z test produced an overabundance of Type I errors. Although Miller (1968) concluded that the S test is robust, other evidence (Fellers, 1972; Games et al., 1972) suggests it is not robust. In all cases the S test showed low power. The Z' statistic was shown by Fellers (1972) to produce erratic, uninterpretable results with small n's, n_i=5. Brown and Forsythe showed it to be robust when distributions were asymmetric and larger n's,

n_i=10, were used. Miller (1968) pointed out that dispersion about a median is not a variance, thus rejecting the technique. Using deviations about the trimmed mean was shown to be nonrobust when distributions were asymmetric by Brown and Forsythe (1974). It appears that none of the Levene statistics has the robustness, power, and flexibility necessary to be recommended as a general variance testing technique.

The Z-variance statistic was shown to be as sensitive to distribution form as the Bartlett test by its developers (Overall and Woodward, 1974). This conclusion was supported by a Monte Carlo study by Levy (1975). Its nonrobustness makes it virtually useless for nonnormal populations.

The jackknife test was offered as a robust and powerful technique by Miller (1968). Layard (1973) concluded that the jackknife test is reasonably robust to moderate deviations from normality. A slight rise in Type I errors for nonnormal populations was shown by Brown and Forsythe (1974). They also noted that with unequal n's the jackknife test produced a rise in P(EI). The effect of unequal n's and possible nonrobustness are areas in which further study is needed for the jackknife test.

The Box test; also referred to as the Bartlett and Kendall test, Scheffé test, and Box-Scheffé test; has been shown to be robust to violations of the assumption of normality in several studies (Fellers, 1972; Games et al., 1972; Gartside, 1972; Layard, 1973; Levy, 1975; Martin and Games, 1975; and Miller, 1968). It has been recommended as a general technique when nonnormality is suspected or when an experimenter has little or no knowledge of distribution form. Unfortunately, it has shown lower power than the jackknife test (Miller, 1968). Miller noted that the Box test is not lagging far behind in power and may be preferred over the

jackknife test for hand computation. The Box test also has shown a rise in error rates with unequal n's (Fellers, 1972; Gartside, 1972). The Bargmann modification appears useful in controlling the effect of unequal n's but at the expense of power. The appropriate sizes for the subsamples when n's are unequal to control the rise in P(EI) and the power of this robust test are areas in which research is needed.

The Box and jackknife tests provide two versatile AOV techniques for analyzing variances. Both easily provide confidence intervals and can be used in complex, multi-factor designs fitting the Bechhofer model. The robustness of one and the effect of unequal n's on both require further study. The present Monte Carlo investigation compares the jackknife and Box tests for normal and nonnormal populations with equal and unequal n's.

2. THE TEST STATISTICS INVOLVED

In this discussion the following notation is used:

k = number of populations involved

s_i^2 = unbiased sample variance from the i^{th} population

d_i = degrees of freedom upon which s_i^2 is based

s_{ij}^2 = unbiased sample variance from the j^{th} subsample from the i^{th} population

v_{ij} = degrees of freedom upon which s_{ij}^2 is based

n_i = size of the sample from the i^{th} population

m_i = number of independent sample variance estimates from the i^{th} population

$Y_{ij} = \log_e (s_{ij}^2)$

p_i = number of subgroups from the i^{th} population

q_{ij} = number of observations in the j^{th} subgroup from the i^{th} population

$Z_{ij} = \log_e (s^2_{ij}) + c_{ij}$

s^2_{i-j} = unbiased sample variance from the i^{th} population with the j^{th} subgroup deleted

$\theta_{ij} = p_i \log_e s^2_i - (p_i - 1) \log_e s^2_{i-j}$

The following test statistics were compared in the study:

2.1 The Box statistic (1953). The n_i observations in each sample are randomly divided into m_i subsamples of $v_{ij} + 1$ observations. If all v_{ij} are equal, the test statistic is:

$$\frac{\Sigma_i m_i (Y_{i.} - Y_{..})^2}{\Sigma_i \Sigma_j (Y_{ij} - Y_{i.})^2} \cdot \frac{\Sigma_i (m_i - 1)}{k - 1}$$

where

$$Y_{i.} = \Sigma_j Y_{ij} / m_i$$
$$Y_{..} = \Sigma_i m_i Y_{i.} / \Sigma_i m_i,$$

and has approximately the F-distribution with $k - 1$ and $\Sigma_i (m_i - 1)$ df.

2.2 The Scheffé modification of the Box test (1959). If the v_{ij} are not all equal, the test statistic is:

$$\frac{\Sigma_i v_i (\eta_{i.} - \eta_{..})^2}{\Sigma_i \Sigma_j v_{ij} (Y_{ij} - \eta_{i.})^2} \cdot \frac{\Sigma_i (m_i - 1)}{k - 1}$$

where

$$\eta_{i.} = \Sigma_j v_{ij} Y_{ij} / v_i$$
$$\eta_{..} = \Sigma_i \Sigma_j v_{ij} Y_{ij} / \Sigma_i \Sigma_j v_{ij}$$
$$v_i = \Sigma_j v_{ij},$$

with $k - 1$ and $\Sigma_i (n_i - 1)$ df.

2.3 The Bargmann modification of the Box test (Gartside, 1972). The variable $Y_{ij} = \log_e (s_{ij}^2)$ is replaced by the variable $Z_{ij} = \log_e (s_{ij}^2) + c_{ij}$. A weighting constant, w_{ij}, is used which is incorporated into the AOV in the usual manner, see Scheffé (1959, pp. 85, 86). The constants c_{ij} and w_{ij} are defined:

$$c_{ij} = 1 / v_{ij} + 1 / (3v_{ij}^2)$$

$$1 / w_{ij} = 2 / v_{ij} + 2 / (v_{ij}^2) + 4 / (3v_{ij}^2),$$

and are used to remove bias and to satisfy better the homoscedasticity assumption of the AOV when n's are unequal. The test statistic is:

$$\frac{\Sigma_i w_i (\eta'_{i.} - \eta'_{..})^2}{\Sigma_i \Sigma_j w_{ij} (Z_{ij} - \eta'_{i.})^2} \cdot \frac{\Sigma_i (n_i - 1)}{k - 1}$$

where

$$\eta'_{i.} = \Sigma_j w_{ij} Z_{ij} / w_i$$

$$\eta'_{..} = \Sigma_i \Sigma_j w_{ij} Z_{ij} / \Sigma_i \Sigma_j w_{ij}$$

$$w_i = \Sigma_j w_{ij},$$

with $k - 1$ and $\Sigma_i (n_i - 1)$ df.

2.4 The jackknife test as developed by Miller (1968) and generalized by Layard (1973). The data are divided into p_i subgroups of q_{ij} observations. The test statistic is:

$$\frac{\Sigma_i p_i (\theta_{i.} - \theta_{..})^2}{\Sigma_i \Sigma_j (\theta_{ij} - \theta_{i.})^2} \cdot \frac{\Sigma_i (p_i - 1)}{k - 1}$$

where

$$\theta_{i.} = \Sigma_j \theta_{ij} / p_i$$

$$\theta_{..} = \Sigma_j p_i \theta_{i.} / \Sigma_i p_i,$$

with $k - 1$ and $\Sigma_i (p_i - 1)$ df.

3. DESIGN OF THE SAMPLING EXPERIMENT

The statistics were compared by empirical simulation using Monte Carlo methods. For each statistic under each experimental condition a simulated analysis was conducted. In each analysis all statistics were computed on the same set of data. A simulated analysis was repeated 1,000 times in four blocks of 250 trials. The number of rejections of the null hypothesis at both the nominal five percent and one percent levels of significance were recorded.

Three distribution forms were used to investigate robustness to nonnormality: normal, N(0,1); moderately leptokurtic, χ^2_4; and extremely leptokurtic, χ^2_2. Populations of 10,000 cases were randomly generated and placed in storage, and population parameters calculated. The N(0,1) population was generated using a random normal number generator developed by Chen (1971). The χ^2_4 and χ^2_2 populations were generated by forming elements of the sum of four squared normal deviates and two squared normal deviates respectively. These normal deviates were generated using the Chen algorithm. The population parameters, μ, σ^2, γ_1, γ_2, are presented in Table 1.

- -

INSERT TABLE 1 ABOUT HERE

- -

Each simulated experiment was conducted on three random samples, representing three treatment groups, with sample sizes of n_i = 7,7,7;

n_i = 17,17,17; n_i = 6,12,18; and n_i = 7,11,17. Sampling was random with replacement. The odd n cases were selected to compare the three Box tests. Equal n's were used to investigate robustness to nonnormality for equal n's and to investigate for small and odd sample sizes the suggestions for selecting subsample sizes for the Box test made by Martin and Games (1975). The unequal n cases were used to investigate the effect of unequal n's combined with heterogeneous within cell variances and robustness to nonnormality for unequal n's.

The null case and three alternatives to the null were investigated. For the non-null conditions, the data of the second and third samples were multiplied by a constant to produce larger variances than in the first sample.

For each sample size condition two sets of subsample sizes were compared for the Box procedures. In the equal n case the two sets were compared to determine which produces the greater power and the effect of the two weighting procedures of Scheffé (1959) and Bargmann (Gartside, 1972) on power. In the unequal n case, v's suggested by the Martin and Games procedure, unequal varying v's associated with unequal n's, were compared with using a constant moderate v for all three samples. For both sample size cases, deleting observations one at a time and two at a time were compared for the jackknife test.

4. RESULTS

The results for the equal n case are presented for $\alpha = .05$ in Table 2. For the three Box procedures V1 refers to the smaller subsample size, v=2

for n_i = 7,7,7 and v=3 for n_i = 17,17,17; and V2 refers to the larger subsample size, v=3 for n_i = 7,7,7 and v=4 for n_i = 17,17,17. For the jackknife tests v is the number of observations being deleted. When the variance ratio is 1:1:1 the empirical error rates have an expected value of .05 if all the assumptions are met. Other variance ratios provide empirical probabilities for points on the power curves. Empirical probabilities of rejection at nonnull points for tests whose error rates deviated significantly at the .05 level from alpha are enclosed by parentheses.

- -

INSERT TABLE 2 ABOUT HERE

- -

When the populations were leptokurtic, the two jackknife tests were not robust. The frequencies of Type I errors were significantly greater than alpha for both population forms. The Bargmann modification with the smaller v (for both n's) was not robust when the population was extremely leptokurtic. All other Box procedures were robust.

When the jackknife tests could be used, i.e., the normal population, they were more powerful than the Box tests, with v=1 more powerful than v=2 regardless of the sample size. For n_i = 7,7,7, the Bargmann modification with v=2 produced the greatest power of the Box tests. The other tests produced approximately equal power. When populations were nonnormal the jackknife tests could not be compared and the Bargmann modification with v=2 was most powerful but it too was not robust for the extremely leptokurtic population. The other tests were approximately equal in power with v=2 slightly more powerful than v=3 and the Scheffé

procedure slightly more powerful than the Box procedure. When n_i = 17,17,17 the results were essentially the same in terms of tests. For the normal population v=4 was more powerful than v=3 but the opposite was true for the leptokurtic populations.

The results for the unequal n cases are presented in Table 3. For the three Box procedures ≠ v refers to using subsample sizes chosen on the basis of the Martin and Games (1975) suggestions, i.e., v=2 for n_1 = 6(7), v=3 for n_2 = 12(11), and v=4 for n_3 = 18(17). Constant v and =v refer to using the same v, v=3 was chosen, for all n's. When n_i = 6,12,18 and v=3 the three Box procedures are identical, hence there is only one entry named Constant v. Empirical powers, enclosed by parentheses because the error rates deviated from alpha, are marked with an asterisk if the deviation was in the conservative direction.

- -

INSERT TABLE 3 ABOUT HERE

- -

The Box test with a constant v was the only test which had no significant deviations from alpha. The Scheffé test with a constant v showed only one significant deviation and the Bargmann modification with unequal v's showed only two deviations but all three were in the conservative direction. A conservative test is not a serious problem for an experimenter unless power is also adversely affected. The Scheffé method was slightly more powerful than the Box method even when the former test was conservative. Unfortunately the Bargmann modification with unequal v's produced much lower power than either the Box or Scheffé methods with even n's and with odd n's when large variances were associated with large

n's. Its power was between the other two tests when large variances were associated with small, odd n's. The other methods, including the jackknife tests, all showed positive deviations from alpha, either as a result of unequal n's, nonnormality, or a combination of the two.

When the assumption of normality was met the Box and Scheffé tests with unequal v's and the jackknife tests were affected by the combination of unequal n's and heterogeneous within cell variances. When the assumption of normality was violated the Box and Scheffé tests maintained approximately the same permissive empirical error rates as with the normal population. This suggests that the tests are robust to nonnormality but not to heterogeneous within cell variances and unequal n's. When the assumption of normality was violated the jackknife tests were even more nonrobust. As the population leptokurtosis increased the empirical error rates also increased showing sensitivity to nonnormality. The empirical error rates were greater than those for the equal n cases suggesting that the jackknife test is also nonrobust to violations of the homogeneity of within cell variance assumption. Only the Box and Scheffé tests with a constant v and the Bargmann modification with $\neq$ v were robust to violations of both assumptions.

5. CONCLUSIONS

At least three forms of the Box test, the original Box test and the Scheffé test when n's are equal, and those two with equal v's and the Bargmann modification with unequal v's when n's are unequal, are robust

to violations of the assumptions of normality and homogeneity of within cell variances when n's are unequal. The Bargmann modification, with equal n's or with equal v's when n's are not equal, was only slightly nonrobust when the population was extremely leptokurtic. It may be considered robust for all but extreme nonnormality.

Using unequal v's when n's are unequal for the Box and Scheffé tests produced an overabundance of Type I errors. When varying v's are used the larger v's and m's are associated with the larger n's. Because of the larger v's the within cell variance is smallest for the largest m and hence largest for the smallest m. This condition is known to produce a permissive bias in the AOV and is an intrinsic outcome as long as larger v's are used with the larger n's. Using a constant v stabilizes these variances. Gartside (1972) had suggested the Bargmann modification as a method to control for heterogeneity of within cell variances in the equal n case. Unfortunately this control is at the expense of power. Both the Box and Scheffé tests (with equal v's) are robust and are more powerful.

The jackknife tests were not robust to violations of the normality assumption for even moderate leptokurtosis. For leptokurtic populations these tests produced P(EI)'s significantly greater than alpha for both equal and unequal n's. For unequal n's the jackknife test is also affected by intrinsic violations of the assumption of homogeneity of within cell variances. For each pseudo-value, the larger the n, the larger will be the number of observations on which the variances are calculated and the more stable the variance estimates will be. Since the p's are directly proportional to the n's, the same permissive condition occurs as with the unmodified Box tests. The jackknife test, without modification, should

only be used with normal populations and equal n's. The jackknife test could be used with a nominal alpha of .01. Data from this study but not presented showed that by using a nominal alpha of .01, it is reasonable to specify that the true risk of a Type I error is approximately .05 or less. The power in this case is approximately equal to that of the Box tests when populations are leptokurtic. An obvious disadvantage of this procedure is the likelihood of confusion on the part of naive users when a nominal alpha differs from the true risk of Type I error.

The Scheffé test, the slightly less powerful Box test, or possibly the Bargmann modification if populations are not extremely nonnormal and more power is desired, can be used as general tests of homogeneity of variance. If care is taken to ensure that the appropriate subsample sizes are used the test is robust and provides satisfactory power. Follow-up comparisons were presented by Games et al. (1972) and the test was extended to the multivariate case by Levy (1974).

REFERENCES

Bartlett, M. S. and Kendall, D. G. The statistical analysis of variance-heterogeneity and the logarithmic transformation. Journal of the Royal Statistical Society, Supplement 8, 1946, 128-138.

Bechhofer, R. E. A multiplicative model for analyzing variances which are affected by several factors. Journal of the American Statistical Association, 1960, 55, 245-264.

Box, G. E. P. Non-normality and tests on variances. Biometrika, 1953, 40, 318-335.

Brown, M. B. and Forsythe, A. B. Robust tests for the equality of variances. Journal of the American Statistical Association, 1974, 69, 364-367.

Chen, E. H. Random normal number generator for 32-bit-word computers. Journal of the American Statistical Association, 1971, 66, 400-403.

Fellers, R. R. The effects of nonnormality and sample size on the robustness of tests of homogeneity of variance. Paper presented to the Northeast Educational Research Association, 1972.

Games, P. A., Winkler, H. B. and Probert, D. A. Robust tests for homogeneity of variance. Educational and Psychological Measurement, 1972, 32, 887-909.

Gartside, P. S. A study of methods for comparing several variances. Journal of the American Statistical Association, 1972, 67, 342-346.

Glass, G. V. and Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs: Prentice-Hall, 1970.

Layard, M. W. J. Robust large-sample tests for homogeneity of variances. Journal of the American Statistical Association, 1973, 68, 195-198.

Levene, H. Robust test for equality of variances. In I. Olkin (ed.) Contributions to probability and statistics. Stanford, Calif.: Stanford University Press, 1960, 278-292.

Levy, K. J. A procedure for testing the equality of p correlated variances. Paper presented to the Psychometric Society, December 1974.

Levy, K. J. An empirical comparison of the Z-variance and Box-Scheffé tests for homgeneity of variance. Psychometrika, 1975, 40, 519-524.

Martin, C. G. and Games, P. A. Selection of subsample sizes for the Bartlett and Kendall test of homogeneity of variance. Paper presented to the American Educational Research Association, April 1975.

Miller, R. G., Jr. Jackknifing variances. Annals of Mathematical Statistics, 1968, 39, 567-582.

Neel, J. H. and Stallings, W. M. A Monte Carlo study of Levene's test of homogeneity of variance: Empirical frequencies of Type I error in normal distributions. Paper presented to the American Educational Research Association, April 1974.

Overall, J. E. and Woodward, J. A. A simple test for heterogeneity of variance in complex factorial designs. Psychometrika, 1974, 39, 311-318.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.

FOOTNOTE

* Charles G. Martin is a Personnel Research Psychologist, Personnel Research and Development Center, U. S. Civil Service Commission, Washington, D. C. 20415. Paul A. Games is Professor, Department of Educational Psychology, The Pennsylvania State University, University Park, PA 16802.
This paper is based on the first author's Ph.D. thesis submitted to The Graduate School, The Pennsylvania State University, 1975. Appreciation is expressed to the College of Education for providing the necessary computer funds.

TABLE 1

Statistical Characteristics of the Generated Populations

Parameters	Normal	Populations: Moderately Leptokurtic χ^2_4	Extremely Leptokurtic χ^2_2
N	10,000	10,000	10,000
μ	-0.0013	3.9896	1.9716
σ^2	1.0040	8.2349	4.0290
σ	1.0020	2.8686	2.0072
γ_1 (skewness)	-0.0299	1.4142	2.0252
γ_2 (kurtosis)	0.0111	3.0903	6.0041

TABLE 2

Empirical Error Rates and Power (α=.05), Equal n's

	n_i = 7,7,7				n_i = 17,17,17			
	Variance Ratio ($\sigma_1^2:\sigma_2^2:\sigma_3^2$)							
Statistic	1:1:1	1:9:17	1:18:35	1:36:71	1:1:1	1:3:5	1:4:7	1:9:17
	A. Normal Distribution, N(0,1)							
Box V1	.048	.289	.475	.565	.043	.424	.578	.852
Scheffé V1	.055	.308	.504	.586	.041	.448	.604	.877
Bargmann V1	.049	.365	.571	.680	.048	.459	.635	.890
Box V2	.049	.313	.445	.577	.049	.501	.651	.915
Scheffé V2	.052	.328	.452	.596	.051	.517	.662	.918
Bargmann V2	.051	.338	.472	.619	.049	.524	.668	.925
Jackknife v=1	.057	.649	.836	.938	.059	.746	.880	.996
Jackknife v=2	.048	.549	.755	.864	.059	.702	.863	.993
	B. Moderately Leptokurtic Distribution, χ_4^2							
Box V1	.040	.209	.335	.476	.041	.298	.439	.773
Scheffé V1	.045	.239	.362	.519	.039	.323	.446	.797
Bargmann V1	.054	.294	.432	.616	.047	.355	.470	.827
Box V2	.042	.208	.336	.466	.047	.318	.453	.821
Scheffé V2	.041	.216	.335	.482	.049	.315	.461	.828
Bargmann V2	.042	.236	.347	.500	.051	.319	.460	.832
Jackknife v=1	.091	(.552)	(.709)	(.818)	.083	(.504)	(.651)	(.909)
Jackknife v=2	.071	(.473)	(.629)	(.768)	.065	(.483)	(.621)	(.901)
	C. Extremely Leptokurtic Distribution, χ_2^2							
Box V1	.058	.208	.311	.410	.043	.232	.321	.615
Scheffé V1	.061	.224	.343	.428	.056	.246	.333	.660
Bargmann V1	.077	(.282)	(.420)	(.520)	.066	(.263)	(.359)	(.695)
Box V2	.049	.181	.262	.342	.043	.248	.333	.673
Scheffé V2	.051	.185	.272	.352	.043	.258	.334	.681
Bargmann V2	.055	.200	.281	.366	.047	.257	.337	.686
Jackknife v=1	.131	(.468)	(.595)	(.701)	.107	(.388)	(.500)	(.764)
Jackknife v=2	.099	(.399)	(.537)	(.633)	.099	(.364)	(.483)	(.743)

TABLE 3

Empirical Error Rates and Power (α=.05), Unequal n's

Statistic	Variance Ratio ($\sigma_1^2:\sigma_2^2:\sigma_3^2$) 1:1:1	1:5:9	1:10:19	1:14:27	1:1:1	9:5:1	19:10:1	27:14:1
			n_i = 6,12,18					
			A. Normal Distribution, N(0,1)					
Constant v	.062	.365	.571	.653	.046	.493	.791	.866
Box ≠ v	.093	(.563)	(.771)	(.850)	.117	(.274)	(.474)	(.570)
Scheffé ≠ v	.088	(.499)	(.696)	(.763)	.092	(.282)	(.552)	(.659)
Bargmann ≠ v	.054	.186	.320	.399	.055	.455	.733	.802
Jackknife v=1	.073	(.603)	(.832)	(.902)	.058	.820	.984	.994
Jackknife v=2	.069	(.489)	(.727)	(.828)	.061	.791	.976	.992
			B. Moderately Leptokurtic Distribution, χ_4^2					
Constant v	.052	.290	.467	.560	.039	.400	.647	.753
Box ≠ v	.096	(.470)	(.669)	(.780)	.089	(.218)	(.398)	(.480)
Scheffé ≠ v	.077	(.419)	(.580)	(.688)	.067	(.218)	(.453)	(.551)
Bargmann ≠ v	.040	.168	.261	.316	.026	(.337)*	(.589)*	(.686)*
Jackknife v=1	.110	(.473)	(.665)	(.755)	.091	(.618)	(.834)	(.877)
Jackknife v=2	.091	(.406)	(.577)	(.676)	.084	(.616)	(.817)	(.862)
			C. Extremely Leptokurtic Distribution, χ_2^2					
Constant v	.045	.233	.357	.452	.060	.264	.495	.609
Box ≠ v	.105	(.421)	(.581)	(.674)	.088	(.168)	(.275)	(.377)
Scheffé ≠ v	.072	(.341)	(.488)	(.568)	.073	(.153)	(.282)	(.405)
Bargmann ≠ v	.028	(.119)*	(.231)*	(.269)*	.041	.232	.399	.517
Jackknife v=1	.142	(.407)	(.526)	(.641)	.136	(.508)	(.668)	(.742)
Jackknife v=2	.133	(.332)	(.467)	(.555)	.133	(.485)	(.652)	(.714)
			n_i = 7,11,17					
			A. Normal Distribution, N(0,1)					
Box = v	.046	.384	.556	.679	.038	.466	.731	.829
Scheffé = v	.044	.400	.584	.718	.035	(.484)*	(.762)*	(.847)*
Bargmann = v	.056	.420	.625	.752	.040	.518	.770	.870
Box ≠ v	.068	(.556)	(.753)	(.844)	.061	.327	.552	.627
Scheffé ≠ v	.059	.470	.667	.778	.053	.366	.634	.721
Bargmann ≠ v	.050	.311	.495	.588	.044	.497	.749	.849
Jackknife v=1	.058	.651	.875	.946	.049	.807	.974	.993
Jackknife v=2	.061	.553	.808	.899	.055	.782	.961	.989
			B. Moderately Leptokurtic Distribution, χ_4^2					
Box = v	.059	.272	.463	.561	.046	.360	.601	.738
Scheffé = v	.058	.286	.490	.597	.055	.386	.625	.772
Bargmann = v	.069	(.314)	(.517)	(.631)	.059	.407	.640	.788
Box ≠ v	.078	(.441)	(.674)	(.751)	.065	(.237)	(.428)	(.554)
Scheffé ≠ v	.065	(.380)	(.564)	(.672)	.057	.262	.498	.638
Bargmann ≠ v	.059	.241	.383	.495	.045	.366	.607	.758
Jackknife v=1	.105	(.514)	(.746)	(.792)	.110	(.627)	(.827)	(.893)
Jackknife v=2	.105	(.458)	(.676)	(.742)	.099	(.606)	(.808)	(.882)
			C. Extremely Leptokurtic Distribution, χ_2^2					
Box = v	.046	.215	.316	.412	.050	.272	.483	.568
Scheffé = v	.054	.237	.346	.446	.046	.293	.504	.595
Bargmann = v	.067	(.273)	(.386)	(.481)	.053	.306	.521	.619
Box ≠ v	.078	(.402)	(.550)	(.647)	.077	(.184)	(.344)	(.397)
Scheffé ≠ v	.064	(.326)	(.440)	(.548)	.066	(.194)	(.390)	(.457)
Bargmann ≠ v	.057	.224	.300	.373	.056	.271	.493	.583
Jackknife v=1	.130	(.431)	(.560)	(.646)	.124	(.479)	(.672)	(.741)
Jackknife v=2	.128	(.373)	(.529)	(.600)	.116	(.458)	(.657)	(.732)

* Test was conservative.